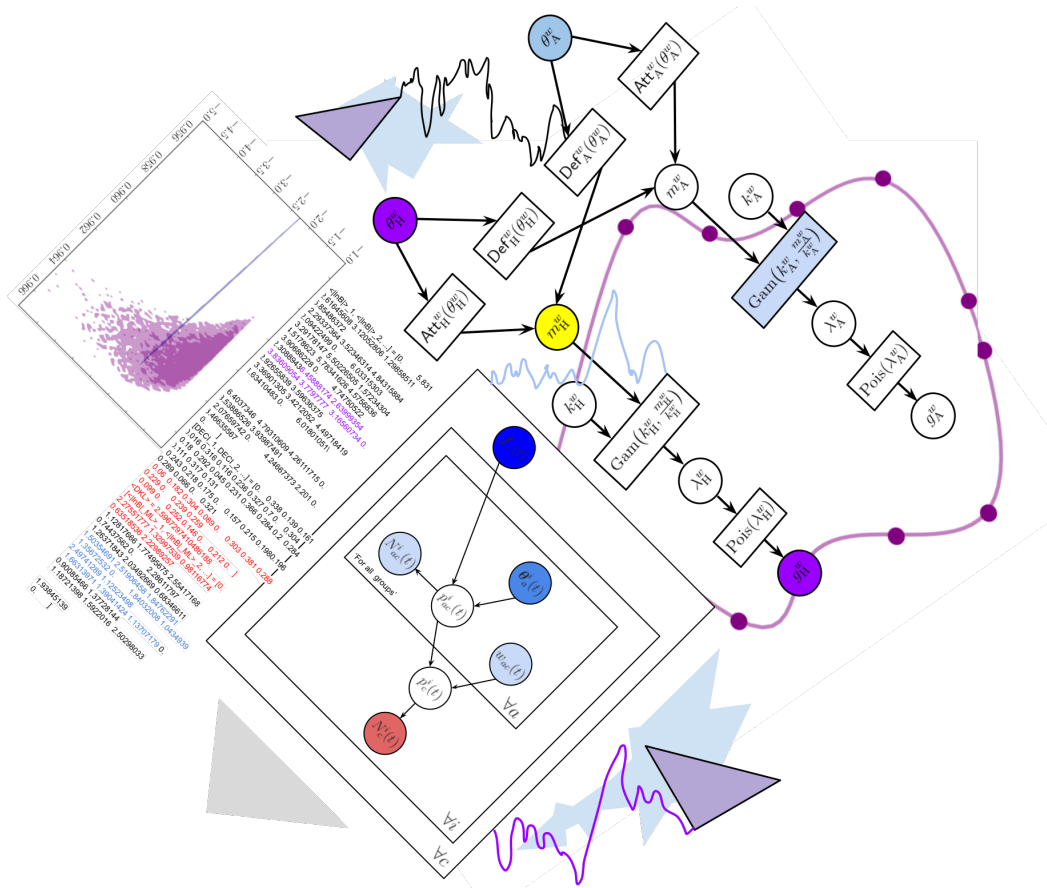


# Diffusing ideas

Adventures with noise and building mathematical toys



Robert J. Hardwick

[umbralcalc@gmail.com](mailto:umbralcalc@gmail.com)

## INTRODUCTION

# 1 | Building the ‘stochadex’

**Concept.** The idea here is

## CORE MATHEMATICAL FORMALISM

Ideally, the stochadex sampler should be designed to try and maintain a balance between performance and flexibility of utilisation (it’s meant to be a general stochastic modelling tool after all!). But before we dive into the software, we need to define the mathematical approach we’re going to take. It seems reasonable to start with writing down the following formula which describes iterating some arbitrary process forward in time (by one finite step) and adding a new row each to some matrices  $S' \rightarrow S$  and  $X' \rightarrow X$

$$(1) \quad X_{t+\delta t}^i = F_{t+\delta t}^i(X', t) + S_{t+\delta t}^i(X', S', t),$$

where to explain things in a bit more detail, let’s define

$\delta t$  as the size of the next timestep

$i$  as an index for the dimensions of the ‘state’ space

$t$  as the current time which may either correspond to the index for some arbitrary stepping process (where  $\delta t = 1$ ), or the explicit time variable for a discrete-time process or some discrete approximation to a continuous-time process

$X$  as the next version of  $X'$  after one timestep (and hence one new row has been added)

$S$  as the next version of  $S'$  (similarly to  $X$  and  $X'$ )

$F_{t+\delta t}^i(X', t)$  as the latest element of some matrix which very generally can depend on  $X'$ ,  $t$  and  $\delta t$

$S_{t+\delta t}^i(X', S', t)$  as the recorded values of a stochastic process defined over the dimensions of the state space indexed by  $i$  which may either vary with continuous sample paths in time or not, and which very generally can depend on  $X'$ ,  $S'$ ,  $t$  and  $\delta t$ .

So the idea here is to iterate the matrices  $X$  and  $S$  forward in time by a row, and use their previous versions ( $X'$  and  $S'$ ) as entire matrix inputs into functions which populate the elements of their latest rows. Note that I’ve avoided using the time variable  $t$  and its step size  $\delta t$  (reducing the formalism to that of an arbitrary stepping process) to try and make the diagram above more clear, but, more generally, these values may be present in the calculations of new matrix rows.

Why go to all this trouble of storing matrix inputs for previous values of the same process? For a large class of stochastic processes this memory of past values is essential to consistently construct the sample paths moving forward. This is true in particular for *non-Markovian* phenomena (see, e.g., [Markov chains](#) to get a sense of their antithesis), where the latest values don’t just depend on the immediately previous ones but can depend on values which occurred much earlier in the process.

## FLAVOURS OF NOISE WITH CONTINUOUS SAMPLE PATHS

For *Wiener process noise*, adopting the [Itô interpretation](#) in this section, we can define  $W_t^i$  is a sample from a Wiener process for each of the state dimensions indexed by  $i$  and our formalism becomes

$$(2) \quad X_{t+\delta t}^i = F_{t+\delta t}^i(X', t) + W_{t+\delta t}^i - W_t^i.$$

Other interpretations of the noise are less immediately compatible with our formalism as it is currently written, e.g., [Stratonovich](#) or others within the  $\alpha$ -family, but it seems less necessary to complicate the details of this section further, so we’ll just cover these extensions at the software implementation level. Note also that we may also allow for correlations between the noises in different dimensions.

For *Geometric Brownian motion noise*, we simply have

$$(3) \quad X_{t+\delta t}^i = F_{t+\delta t}^i(X', t) + \textcolor{red}{X}_t^i(W_{t+\delta t}^i - W_t^i).$$

And say, e.g., *fractional Brownian motion noise*, where  $B_t^i(H_i)$  is a sample from a fractional Brownian motion process with Hurst exponent  $H_i$  for each of the state dimensions indexed by  $i$ , we simply substitute again

$$(4) \quad X_{t+\delta t}^i = F_{t+\delta t}^i(X', t) + \textcolor{red}{B}_{t+\delta t}^i(H_i) - B_t^i(H_i).$$

*Generalised continuous noises* would take the form

$$(5) \quad X_{t+\delta t}^i = F_{t+\delta t}^i(X', t) + \textcolor{red}{g}_{t+\delta t}^i(X', W_{t+\delta t}^i - W_t^i, \dots),$$

where  $g_{t+\delta t}^i(X', W_{t+\delta t}^i - W_t^i, \dots)$  is some continuous function of its arguments which can be expanded out with [Itô's Lemma](#).

## FLAVOURS OF NOISE WITH DISCONTINUOUS SAMPLE PATHS

*Jump process noises* generally could take the form

$$(6) \quad X_{t+\delta t}^i = F_{t+\delta t}^i(X', t) + \textcolor{red}{J}_{t+\delta t}^i(X', \dots),$$

where  $J_{t+\delta t}^i(X', \dots)$  are samples from some arbitrary jump process (e.g., compound Poisson) which could generally depend on a variety of inputs, including  $X'$ .

*Poisson process noises* would generally take the form

$$(7) \quad X_{t+\delta t}^i = F_{t+\delta t}^i(X', t) + \textcolor{red}{N}_{t+\delta t}^i(\lambda_i) - N_t^i(\lambda_i),$$

where  $N_t^i(\lambda_i)$  is a sample from a Poisson process with rate  $\lambda_i$  for each of the state dimensions indexed by  $i$ . Note that we may also allow for correlations between the noises in different dimensions.

*Time-inhomogeneous Poisson process noises* would generally take the form

$$(8) \quad X_{t+\delta t}^i = F_{t+\delta t}^i(X', t) + \textcolor{red}{N}_{t+\delta t}^i(\lambda_{t+\delta t}^i) - N_t^i(\lambda_t^i),$$

where  $\lambda_t^i$  is a deterministically-varying rate for each of the state dimensions indexed by  $i$ .

*Cox (doubly-stochastic) process noises* would generally take the form

$$(9) \quad X_{t+\delta t}^i = F_{t+\delta t}^i(X', t) + \textcolor{red}{N}_{t+\delta t}^i(\Lambda_{t+\delta t}^i) - N_t^i(\Lambda_t^i),$$

where the rate  $\Lambda_t^i$  is now a sample from some continuous-time stochastic process (in the positive-only domain) for each of the state dimensions indexed by  $i$ .

*Self-exciting process noises* would generally take the form

$$(10) \quad X_{t+\delta t}^i = F_{t+\delta t}^i(X', t) + \textcolor{red}{N}_{t+\delta t}^i[\mathcal{I}_{t+\delta t}^i(N', \dots)] - N_t^i[\mathcal{I}_t^i(N'', \dots)],$$

where the stochastic rate  $\mathcal{I}_t^i(N', \dots)$  now depends on the history of  $N'$  explicitly (amongst other potential inputs - see, e.g., [Hawkes processes](#)) for each of the state dimensions indexed by  $i$ .

*Generalised probabilistic discrete state transitions* would take the form

$$(11) \quad X_{t+\delta t}^i = /F_{t+\delta t}^i(X', t) + \textcolor{red}{T}_{t+\delta t}^i(X'),$$

where  $T_{t+\delta t}^i(X')$  is a generator of the next state to occupy. This generator uses the current state transition probabilities (which are generally conditional on  $X'$ ) at each new step.

## SUMMARY OF DESIRABLE FEATURES

- using the learnings from the previous sections looking at specific example processes
- above formalism is so general that it can do anything - so while it shall serve as a useful guide and reference point, it would be good here to go through more of the specific desirable components we want to have access to in the software itself
- it might not always be convenient to have the windowed histories stored as  $S$  but some other varying quantity which can be used to construct  $S$ ? take fractional brownian motion as an example of this! hence, need to provide more possible input histories into  $S$
- want the timestep to have either exponentially-sampled lengths or fixed lengths in time
- formalism already isn't explicit about the choice of deterministic integrator in time
- but also want to be able to choose the stochastic integrator in continuous processes (Itô or Stratonovich?)
- enable correlated noise terms at the sample generator level
- configurable setup of simulations with just yamls + a single .go file defining the terms

Test cite [1]

## DESIGN CHOICES FOR THE STOCHADEX

### REFERENCES

- [1] E. Dimastrogiovanni, M. Fasiello, R. J. Hardwick, H. Assadullahi, K. Koyama and D. Wands, *Non-Gaussianity from Axion-Gauge Fields Interactions during Inflation*, [1806.05474](#).